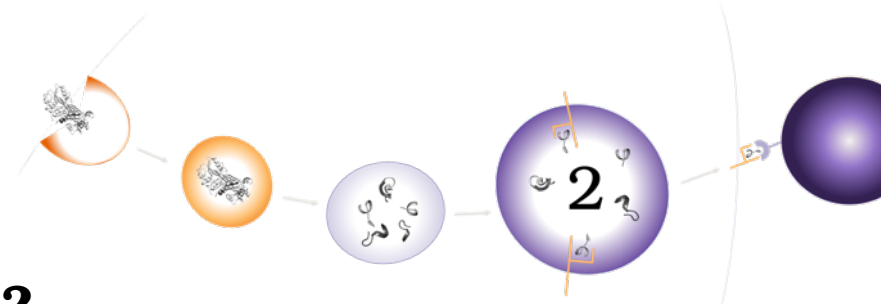
Bioorthogonal antigens as tool for investigation of antigen processing and presentation

Pieper Pournara, L.



Chapter 2

Introduction

Humans are exposed to a wide variety of pathogens. These are usually dealt with efficiently by the cells and proteins of the innate immune system (IS). Through phagocytosis and the controlled activation of anti-microbial mechanisms, most potential pathogens are killed and cleared away before the onset of symptoms. However, due to the evolutionary ancient nature of this response, many pathogens have evolved strategies to circumvent these killing mechanisms. For example, the intracellular pathogens *Chlamydia pneumoniae* and *Mycobacterium tuberculosis* (MTB) can persist for years inside vacuoles within the phagocytic cells of the IS[6–8].

Vertebrates have therefore evolved a second 'layer' of the IS: the adaptive immune system[9–11]. This part of the immune response can respond to specific pathogenic threats by activating cells that have receptors capable of recognizing molecules that are unique to specific pathogenic species. Their activation allows them to orchestrate the most effective immune responses against that species[12]. The adaptive immune cell is also key in regulating tolerance against certain antigens[13–15]. The tolerogenic activation of adaptive immune cells recognizing potential allergens, commensals, and self-antigens, leads to a potent dampening of the immune response against these species. Both these roles make the control over the adaptive immune response of a key pharmaceutical interest[16].

There are two classes of adaptive immune cells, T lymphocytes (T cells) and B lymphocytes (B cells)[17], that have the ability to change their genome at a specific locus[18]. The latter recognize antigens through their B cell receptor[19–21], leading to the production of antibodies that target the antigen for elimination[19,22]. The biochemistry behind the recognition of antigens by T cells is more subtle. These cells recognize peptides derived from the antigen, rather than the intact antigen[23–25]. Furthermore, these peptides have to be presented on host-proteins called major histocompatibility complexes (MHCs) for the T cells to recognize them[26].

The major types of antigen presenting cells (APCs) are dendritic cells (DC), macrophages and B cells[27–29]. DCs are most efficient in this task and often referred to as professional APCs[30,31]. DCs are tissue-resident myeloid derived cells, characterized by a high capacity of antigen uptake. In their immature state, however, they possess a low T cell stimulatory capacity[32]. When they encounter antigen together with a danger or damage signal, they enter a maturation program, which makes the cells capable of migrating through the lymphatics towards (secondary) lymphoid tissue (such as lymph

nodes (LN))[33]. At the same time their antigen uptake capacity is reduced and their cell surface levels of MHCs increased, together with an up regulation of co-stimulatory molecules (which are required for the pro-inflammatory activation of T cells specific for the antigen)[34]. By the time they arrive in the LN, they have thus converted from a macrophage-like cell to a cell highly equipped for activating T cells. DCs can therefore be viewed as the sentinels of the IS and the study of their handling of antigenic cargo can aid, for example, the design of new prophylactic vaccines, therapeutic vaccines and cancer immunotherapies[27,35].

The key processes that underpin this T cell activation by DCs are the uptake of antigen, its proteolysis into peptide fragments and the loading of these peptide fragments onto MHCs. These peptide-MHC (pMHC) complexes are sampled by the 10^7 different T cells an average human possesses[36]. Two structurally distinct MHC-alleles are known, namely, MHC class-I (MHC-I) and MHC class-II (MHC-II)[37,38]. They differ in the source of the presented peptides (endogenous versus exogenous) as well as the subtype of T cell they interact with: MHC-I present peptides from a cytosolic origin leading to the activation of CD8 (cytotoxic) T cells (CTL)[39]. MHC-II present peptides from exogenously taken up material, leading to the activation of $CD4^+$ (helper) T cells[40,41].

The numbers behind antigen processing and presentation are astonishing: an approximate 10^{12} to 10^{15} different peptides can bind MHC from the human proteome, yet only 10,000 to 500,000 MHC-I[41] and a similar number of MHC-II molecules are present per APC[42,43]. The MHC molecules have a turn over time of ~ 24 h and can display about 4,000 new peptides on the cell surface each hour[44]. Therefore, the chance of a T cell encountering one of its cognate peptides in high copy numbers is very low. Active recruitment - through the use of chemokines - is thus employed by the DC to maximize this chance[45]. All circulating T cells will pass through a LN containing activated DCs within 12-24 h[11]. This does not negate the low copy numbers of specific pMHCs, especially against a background of other common pMHCs[46]. This is a problem, which is further compounded by the low K_D-values of a T cell receptor for a given pMHC[24,47]. This means that subtle differences in binding parameters must be able to yield distinct immunological outcomes[48]. The underlying biology of this phenomenon remains to be elucidated, but it seems essential that this process can happen very fast (to ensure timely clearance of pathogens or aberrant cells) and with a high degree of specificity, as a misdirected T cell response can have devastating consequences: auto-immune diseases, such as type-1 diabetes (T1D)[49], multiple sclerosis[50] or rheumatoid arthritis (RA)[51] show[48].

All the mentioned factors - the number of MHCs per APC, the number of specific pMHCs per cell and the time of turnover and cellular concurrence - have to be considered for the further development and improvement of vaccines. One area that is under particular scrutiny at present, is the use of therapeutic vaccination in the fight against certain types of cancer[52–54]. Here a patient with an established tumour is vaccinated against antigens derived from this tumour, in an attempt to restart the anti-tumour immune response. The use of vaccines against cancer is beginning to make its mark on the treatment of cancer[55]. In particular, the recent development of so-called neo-epitope vaccines is looking very promising[56–58] (see Figure 2.1).

Neo-epitopes are T cell antigen epitopes that are newly created by mutations in the coding regions of the tumour DNA[54,59,60]. In certain cancer types, mutations can appear in remarkable numbers. A study of pre-malignant colorectal cancer sample analysis demonstrated a total number of 11,000 mutations in a single cancerogenic

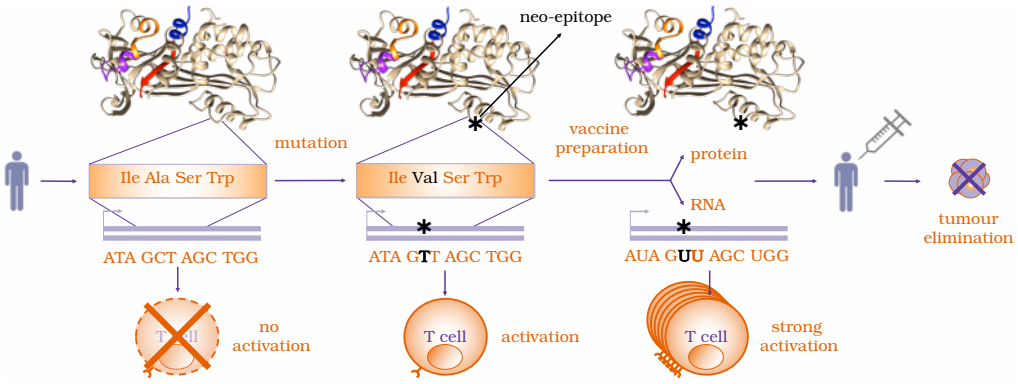


Figure 2.1: Neo-epitope emergence in tumours and their use for anti-cancer vaccine development. Neo-epitopes emerge driven by the high mutation rate in tumours. Although there was no T cell response evoked by a tumour derived protein before, a mutation can cause a neo-epitope to appear and readily lead to a potent new anti-tumour immune response. This phenomenon is then translated into a patient-specific protein- or RNA-based vaccine, which can be administered to the patient in return (re)activating their immune system.

cell[61]. More recently, an estimate of the neo-antigen repertoire in human melanoma was published. The estimation shows approximately 150 amino acid changing mutations per individual[62]. However, the adaptive immune response against these neo-epitopes is often absent or has stalled during tumour emergence[62–64].

In an attempt to (re)initiate this anti-neoepitope immune response, two similar approaches were recently reported. Both made use of personalized cancer vaccines based on the individual mutations in a person's tumour genome[58,65]. First the transcriptome of both the tumour and healthy cells were sequenced and the tumour-specific exome mutations were identified. These were then classed based on expression and MHC-binding ability. Here the two approaches diverged. While Wu and co-workers[66] made a peptide vaccine based on these tumours[66], Sahin and co-workers made an RNA-based vaccine[58], which were then used to immunize their patients. Both approaches (in phase I) looked promising, with Wu *et al.* reporting vaccine-induced poly-functional $CD4^+$- and $CD8a^+$-T cells against neo-epitopes in all vaccinated stage IV melanoma patients[66,67] and robust *de novo* responses against multiple neo-antigens in two out of eight patients with glioblastoma[68]. The study of Sahin *et al.* included 13 patients with stage III or IV melanoma testing a poly neo-epitope vaccine[58]. All patients developed a response against a minimum of 3 neo-epitopes[58,67].

One striking aspect of the two studies above was the inefficiency of the vaccine design: only 1 in 10 potential epitopes elicit an $CD8a^+$ T cell response and most that did elicit an immune response produced an unexpected $CD4^+$ T cell response; the effect(s) of this response, however, remain to be elucidated[69]. One potential reason was that only the strength of a peptide binding an MHC and its expression levels were taken into account for epitope selection. Processing and routing properties were not considered. The likely reason for this is that there are at present no tools, that can reproducibly assess these parameters. The development of new tools for this part of the T cell activation cascade is thus needed. This thesis discusses the development of new methods to better map antigen processing by DCs.

2.1 Mechanisms of antigen processing

Antigen presentation consists of three key events: antigen uptake, proteolysis of the proteinaceous content of the antigen, and loading/presentation of the generated peptides onto MHC-complexes. There are different ways antigen can be taken up: clathrin-mediated endocytosis[70], phagocytosis[71] and macropinocytosis[72], and each of these methods affects the efficiency of downstream antigen presentation[73]. The process of peptide loading is also well controlled, tightly regulated and has been reviewed in detail[74–77]. Phagocytosis targets antigen to a compartment that is highly proficient in antigen presentation and cross-presentation, whereas certain receptors can target antigens to compartments in which one of these forms of presentation is favoured. MHC-II loading is achieved by peptide exchange mediated by the protein human leukocyte antigen-DM (HLA-DM), or by the peptide loading complex (PLC) for MHC-I[78]. In the remainder of the chapter, proteolysis and its role in peptide generation for MHC-loading will be described. The production of peptides from the taken up antigenic material is achieved through proteolysis: the material taken up by the APC is routed to the endo-lysosomal system where the resident proteases effect this degradation. Besides the obvious factors such as nature and abundance of these proteases themselves, their trafficking and maturation, the process of proteolysis is influenced by a plethora of additional factors. These factors include the aforementioned uptake[39], but also the type of APC[79], the triggering of danger receptors on the APC[80], the amount of oxidative burst - the generation of reactive oxygen species, regulating phagosomal pH - that was initiated during the initial uptake[81], to name but a few. The antigen itself also affects the speed and efficiency of the process: the position of a given peptide within the antigen[82] appears to determine the efficiency of antigen presentation[83], as do the physio-chemical properties of the antigen (e.g., protein size, structural tightness[84]), or the presence of post-translational modifications (PTM)[73,85–87]. Studying and predicting the efficiency by which these processes occur is of prime importance to improve the design of vaccines and other immunotherapies. However, the degradation of antigen is hard to study due to this co-dependence of multiple factors, both within the antigen, as within the APC.

2.1.1 Endo-lysosomal pathway

The endo-lysosomal system is a sub-cellular vesicular system in which external and, through autophagy, internal material is degraded. The endo-lysosomal compartment is therefore a key site for the production of peptides from proteins for MHC-loading[88]. The endo-lysosomal compartments alone host more than 50 hydrolytic enzymes, including a very diverse set of proteases. These can be divided on their active site mechanism and four catalytic types have been found: aspartic, cysteine, metallo and serine proteases. They also have been grouped into cathepsin and non-cathepsin proteases, but this bears no relation to their mechanism of hydrolysis: cathepsins include aspartate, serine proteases and papain-like cysteine proteases (PLCP)[89]. Other than these, the main endo-lysosomal proteases that have been identified in antigen processing are asparagine endopeptidase (AEP, legumain)[90], and the furin family of proteases[91]. It is of prime biological importance that these enzymes are only active within the endo-lysosomal system. Aberrant activation of these promiscuous enzymes in other organelles would lead to the unwarranted destruction of host proteins and biological deregulation. Extracellular cathepsin activity has, for example, been implicated as a key factor in the metastatic

ability of cancer[92]. The activity of the endo-lysosomal proteases is regulated on two levels. First of all, they are all produced as inactive pro-enzymes (zymogens) [93] that are rapidly routed to the endo-lysosomal compartment[94]. Here, they are activated using the unique biology of this compartment: activation of the proteases is, partly, regulated by the acidity of the organelle (lysosomal maturation involves the gradual lowering of the pH within the compartment[95] from neutral to as low as a pH of 4.0. Early endosomes have a pH between 6.0-6.5[96]). Along the endo-lysosomal pathway the pH is slowly decreasing down to pH 4.9-6.0 in late endosomes[97]. The mature lysosomes, have a pH of 4.0-5.0[98]. Most of the endo-lysosomal proteases (with an exception for cathepsin B[99]), have pH activity optima in the acidic pH range, rendering them nearly inactive at neutral pH values. The endo-lysosomal pathway is not homogeneous, conditions such as pH and oxidative potential vary between early and late compartments. The activity of the various proteases found in this pathway also varies, due to these changes in pH. Some are more active near neutral pH, whereas others are more active at acidic pH. Their functional interplay (the fact that they activate and degrade one another) adds a further layer of complexity. This results in a complex network of protease activities. It is very important, for the reason already mentioned earlier in this Section, that the proteases should be activated only within the endosomes or lysosomes, where they are needed[100]. The zymogen consists of a signal peptide of 10-20 amino acids[100], which is removed in the endoplasmic reticulum (ER) and N-glycans with mannose-6-phosphate (M6P) moiety are added to target the pro-enzyme to the endo-lysosome[101]. It follows a pro-sequence of 38-250 amino acids that assists the correct folding[100]. Only in the endo-lysosome the pro-sequence is cleaved off for final activation. This can happen only through the decrease of pH in the endo-lysosome compared to the cytosol, with the help of glycosaminoglycans (GAG)[102], by (auto-)catalytic cleavage or through other present proteases, which are already active[103–105].

2.1.2 The role of endo-lysosomal proteases in antigen presentation

The selectivity of endo-lysosomal proteases towards their substrates is rather low. The amino acids in and around the cleavage site offer some selectivity[106]. In Figure 2.2 a scheme of the typical annotation of the binding site is depicted, whereas the scissile bond lies between P and P'.

For the processing of antigens, this means that the process appears rather chaotic; with different proteases having been implicated as the key protease for specific antigens[109,110]. This is further complicated by the fact that the scissile efficiency can also be influenced by PTM of the antigen, such as citrullination and glycosylation[111]. Some of these PTMs, such as the glycans are themselves also enzymatically hydrolysed in the endo-lysosomal pathway. This adds yet another layer of complexity. The same protease families are also important for the activation of the MHC-II receptors in late endosomes by degrading the invariant chain[112,113], as well as for the activation of certain Toll-like receptors (TLR) that affect the co-stimulatory signalling also needed for T cell activation[114]. To give but one example, AEP is an endo-lysosomal of high substrate selectivity. It cleaves only after asparagine (Asn, N) residues. This protease was shown to be involved e.g., in the destructive processing of very specific epitopes containing asparagines, such as the auto-antigen from myelin basic protein (MBP)[115] as well as specific epitopes from model vaccine proteins, such as the C-fragment of tetanus toxin (TTcF)[90]. However, a repeat of some of these processing studies *in vivo* in a knock-out

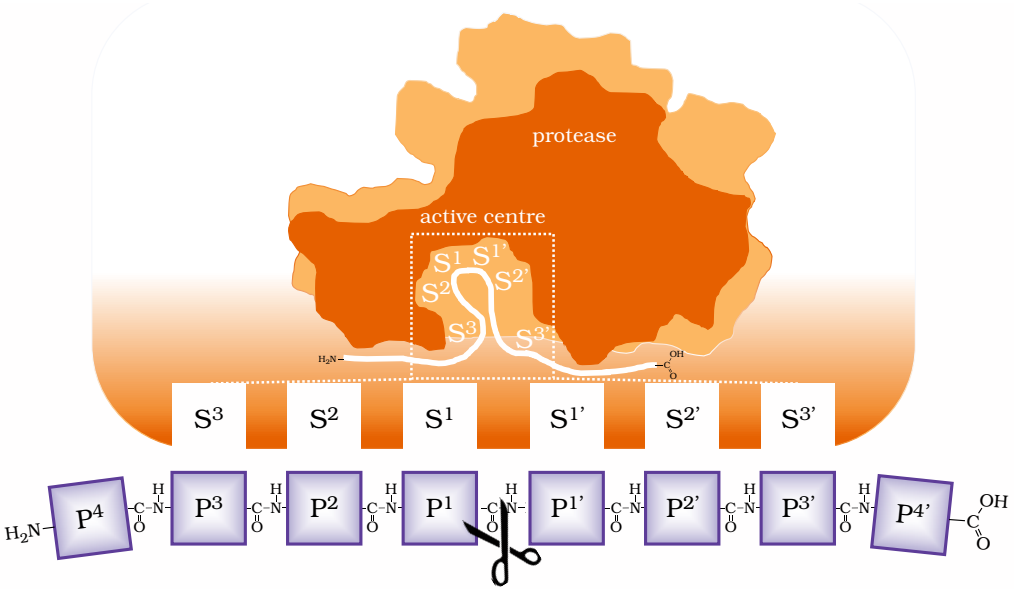


Figure 2.2: Annotation of binding site and amino acid residues for enzymatic active center unit. The peptide bond is cleaved between position P1 and P1' of the amino acid sequence. The bond between these two positions is called scissile bond. Source: Adapted from [107,108].

mouse showed a complete abrogation of the key role of this enzyme observed earlier *in vitro*, suggesting the presence of compensatory mechanisms [116,117]. These differences in proteolysis have been shown *in vivo* using stabilized and destabilized antigens. These showed drastically different immune responses depending on protein stability [118,119].

2.2 Methods of studying antigen processing

The combination of the functional complexity of the protease network and the wide-ranging effects on antigen presentation efficiency, have made it difficult to determine precise details about many aspects of the pathway to date. Extensive research has gone into developing new tools for studying antigen presentation with approaches from various angles. The quantification of peptides appearing after processing has been attempted; through the use of T cells, but also mass spectrometric analysis (MS)-based identification approaches [120]. The subcellular routing of the antigen is also studied using e.g. detectable versions of antigens [121], and the use of inhibitors that knock out specific aspects of the antigen routing pathway [121,122]. This section will describe some of the key assays that were developed to study the route (and associated rates) of antigen presentation.

2.2.1 *In vitro* and intracellular degradation assays

The processing rate for an antigen is most easily studied *in vitro*, using either isolated lysosomes from APCs or recombinant lysosomal proteases [118,119]. Isolated lysosomes from different sources, such as primary isolates or cultured cell lines, have been used

for example by Delamarre *et al.* to show that the structural destabilization of an antigen caused a rapid increase in its rate of proteolytic degradation, which then translated *in vivo* to a marked decrease in the ability of a vaccine based on this destabilized protein to activate a T cell response[118]. The work flow was as follows: recombinant or isolated proteins were either enzymatically (in case of RNase-A) or chemically (in case of horseradish peroxidase (HRP)) destabilized. Both versions were incubated with isolated lysosomes for different time periods before denaturation and sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis to separate the breakdown fragments by size[119]. Alternatively, the resulting breakdown products can be used for loading onto APCs, in which uptake and processing have been eliminated[123]. Fully cell-free systems, for example based on only the essential components MHC-II, cathepsins and HLA-DM have also been developed[124]. Studying the degradation of antigen in cells is more complex: the large amount of protein present in the cell precludes the analysis by SDS-PAGE, see Figure 2.3).

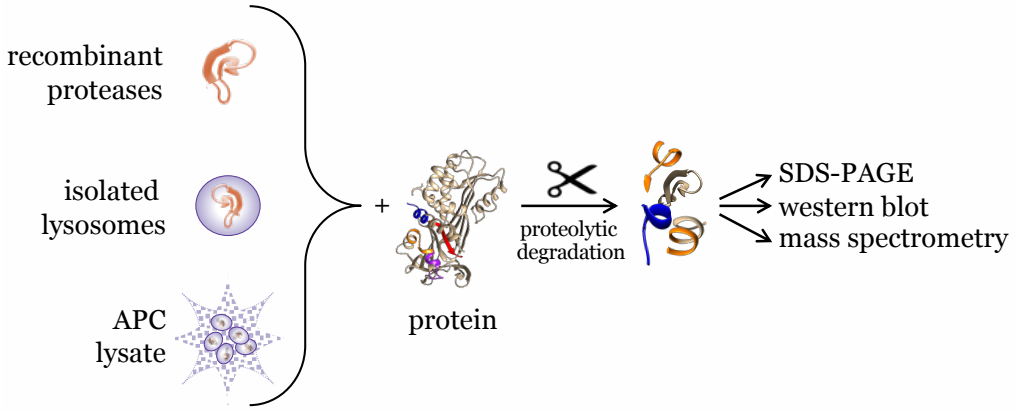


Figure 2.3: *In vitro* and intracellular protein degradation. Proteolytic degradation *in vitro* or intracellular can be done in different ways by recombinant proteases, by extracted lysosomes from primary tissues or cultured cells or by feeding protein to APCs and extract the degraded protein peptides from lysate. Peptide analysis can be done for example by sodium dodecyl sulphate-polyacrylamide gel electrophoresis, western blot or mass spectrometric analysis.

Fluorophore-modified proteins have proven to be valuable reagents for immunological research on antigen uptake and routing studies[121]. Classically, these proteins have been used e.g., to track and trace cells that have taken up such fluorescent proteins through micropinocytosis *in vitro* as well as *in vivo*[125–128] or to detect protein uptake within cell type(s) of interest[120,129–131]. Extensive studies have been performed using fluorescent proteins, but over time more and more intracellular studies have been performed with a focus on questions, such as routing of antigen and their breakdown kinetics within antigen presenting cells[132,133]. Looking at these studies from a chemical perspective raises questions in terms of the chemical properties of fluorophore-modified proteins. Individual studies have raised issues concerning, for example, the use of commercially available quantum dot-ligated bioconjugates[134,135]. The concerns are discussed in more detail in the context of the following chapters.

2.2.2 T cell assays

Most studies of T cell activation have been performed using T cells specific for model antigens. In these experiments, the activation of these cells upon stimulation by a model antigen-loaded APC are used in view of more direct measurements [136]. With different methods different parameters can be measured. For example, the production and excretion of certain cytokines can be measured as quantification of activation, e.g. IFN-γ or even cytokine profiling [137], or the proliferation of T cells themselves [138,139].

Most T cell assays are co-culture assays with two types of cells. As a base APCs are used for the presentation of antigens, such as primary bone marrow-derived dendritic cells (BMDCs), total splenocytes or a DC cell line, for example DC2.4 [140], D1 [141] or MuTu1940 [142]. Most of these cells are from murine or human origin. The antigens of interest are incubated with the APCs for a given amount of time and excess is removed. Optionally, the APCs can be chemically fixed to immobilize the cell membrane and therefore stabilize the amount of peptide presenting MHC-complexes on the cell surface. Then, as the second cell type, primary or immortalised T cells are added bearing T cell receptors (TCR) for the antigen at hand with known TCR specificity. Subsequently, different readout parameters can be measured depending on the need. The time of co-culture depends on the readout parameter. For example, the stimulation of cytokine production or the proliferation of activated T cells take different amounts of time to surpass a measurable detection limit and therefore determine the incubation times respectively. In case of primary T cells cytokine levels are commonly determined using cytokine specific enzyme-linked immunosorbent assay (ELISA) [143], enzyme-linked immune absorbent spot (ELISpot) assay [144] or intracellular cytokine staining (ICS) [145,146].

Immortalized T cell lines are also used often. They can, for example, be transfected with a reporter gene enabling colorimetric readout via the expression of a reporter enzyme (e.g., β-galactosidase (β-gal) or luciferase). The enzyme expression is coupled to the promoter of an activation-related cytokine, usually, for T cells, this is NFAT. This enzyme can then be used to specifically quantify cleavage of a coloured substrate via colour change without (too much) interference of endogenous levels of this enzyme [147,148]. T cell proliferation for primary and T cell lines can be measured using either radioactivity or fluorescence. The principle of [^{3}H]-thymidine incorporation is based on the marking of cells with a radioactive nucleoside, [^{3}H]-thymidine, which is incorporated in the deoxyribonucleic acid (DNA) of the cells and which is then diluted with a factor 2 with every cell division cycle. This leads to a continuous dilution of the radioactive signal, which can then be measured. Quantity and quality of the proliferation can thus be determined. Nowadays, a less hazardous version using fluorescence, i.e., CFSE staining, is utilized, which is cell permeable and couples to intracellular lysine residues or other amines. The measurement is following the same principle of a factor 2 dilution per cell division cycle. As a readout, flow cytometry can be used [149,150]. For T cell assays the needed T cells have to be generated. For that animals are either vaccinated with the respective antigen to harvest antigen specific T cells or T cells from specific mouse strains are necessary. The alternative option are T cell lines created specific for certain antigens by fusion of an antigen specific primary T cell with an immortal thymoma cell line [151] (for details refer to van Endert *et al.* 2013 [152,153]). A large number of antigen specific T cell lines are meanwhile available for different antigens [154,155] e.g., chicken egg albumin (B3Z T cell) [156] or the herpes simplex virus-1 (HSV-1) glycoprotein B (HSV GP), HSV2.3, epitope 498-505, SSIEFARL, SL-8) [154,155]. Other common pathogens or

antigens for which antigen-specific T cell lines are available are human papillomavirus (HPV), cytomegalovirus (CMV), the influenza virus, bacterial tetanus toxin or MBP in the context of experimental autoimmune encephalomyelitis (EAE).

2.2.3 Mass spectrometry assays

A more recent approach that allows the determination of peptide levels bound in MHC molecules without the need for T cells is based on the use of MS [157]. The aim of this method is to provide an unambiguous and full identification of the MHC-presented peptides [158,159].

The workflow of the approach starts with the immune-precipitation of the MHC-complexes [160], or by eluting the MHC-bound peptides from the surface of an APC [161]. The MHC-bound peptidome is then eluted and analysed by a variety of state-of-the-art MS techniques that allow the identification of sequence determination of the bound peptides [159]. Immunoprecipitation of MHC-complexes with bound peptide can be done with cultured cells or cells extracted from primary tissue using immunoaffinity purification [162,163]. For example, in case of chronic lymphocytic leukemia (CLL), Kowalewski *et al.* reported to have found tumour associated antigens, which are frequent and exclusive of the HLA-ligandome of leukemic cells [164]. This was done by first determining frequencies of representation in CLL ligandomes correlating with the frequencies of immune recognition by patient T cells and then retrospectively analysing patient survival, which revealed benefits for patients displaying immune responses to these antigens. There are, however, difficulties with this approach. One is the very large cell numbers currently required to obtain good depth of coverage for the identification of peptides. This is caused by a limited reliability of the results of genome and transcriptome sequencing. Especially for tumour cells are always very heterogeneous in their transcriptome [59,60,165]. Another difficulty is the detection of rare peptides, particularly for primary samples. The assessment of the relevance of these rare peptides thus remains hard.

2.3 Summary of limitations of (above) methods

All of the above methods have their limitations. T cells - the most often used detecting agent - have specificity for particular peptides, and following these is not possible, as their chemical modification renders them non-cognate to their target T cell. Even following the protein during the degradative process *in cellulo* is not fully possible at present. The closest approach are fluorophore modified reagents. These allow the tracking of the antigen, but also alter is physical properties. For example, the modification of lysines with fluorophores removes positive charges from the protein surface, whilst at the same time introducing large, lipophilic fluorophore moieties at these sites (see Figure 2.4).

It is here hypothesized that these variations in structure, charge and lipophilicity alter the rate and efficiency of antigen proteolysis and presentation, and any visual data obtained from these reagents may thus not be representative of the wildtype antigen. It is therefore necessary to develop new approaches that allow the tracking of the antigen without this concomitant change in properties.

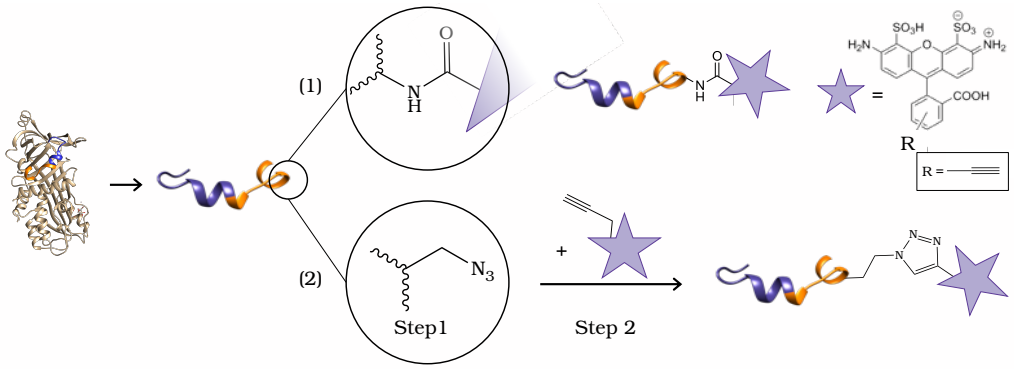


Figure 2.4: One and two-step principle of antigenic peptide modification. Antigenic peptide pre-modified with (1) fluorophore e.g., AlexaFluor, compared to (2) the same peptide modified with an azide containing amino acid being labelled in 2 steps with an alkyne functionalized fluorophore.

2.4 Bioorthogonal chemistry

One area of chemistry that could be of potential use to solve this conundrum are the family of bioorthogonal ligation reactions. These are reactions between two (or more) reaction partners that can be selectively executed in biological samples; with minimal cross-reactivity with the chemical functionalities found herein.

2.4.1 Bioorthogonal reactions

The first example of a chemical ligation that could be selectively performed in a biological medium was a Staudinger-like ligation reported by Bertozzi and co-workers [166,167]. Here, an azide-functionality was metabolically incorporated into the glycoproteome of a cell, through the addition of azide-modified metabolic precursors of sialic acids. These azides, after conversion to the aforementioned sialic acids and incorporated in the glycoproteome of the cell, could then be selectively reacted on the live cell using a modification of the Staudinger reduction reaction [168] (see Figure 2.5).

The phospho-aza-ylide intermediate, which in the Staudinger reduction reaction rearranges to the iminophosphorane and then is hydrolysed to give the amine, is trapped by an intramolecular electrophilic trap to yield an amide ligation product. If the phosphine is also modified with a detectable group (such as biotin), then the amount of azide on the cell can be quantified using this approach. After the introduction of this concept, many new bioorthogonal reactions were developed, as well as new methods for incorporating bioorthogonal reaction partners into biomolecules. The first of these was the copper-catalysed alkyne-azide [3+2]-cycloaddition reaction (CuAAC), which is the reaction between an azide and an alkyne, catalysed by Cu(I). This reaction has very fast kinetics (~ 10-200 $M^{-1}s^{-1}$), but is limited in its 'bioorthogonality' due to the toxicity of the copper-catalyst used (Reaction a, Figure 2.6).

Bertozzi and co-workers developed a variant of this reaction, which did not require the copper catalyst [169,170]. Instead of a linear alkyne, a strained alkyne was used as reaction partner for the azide. The strain on the alkyne renders it more reactive, pre-

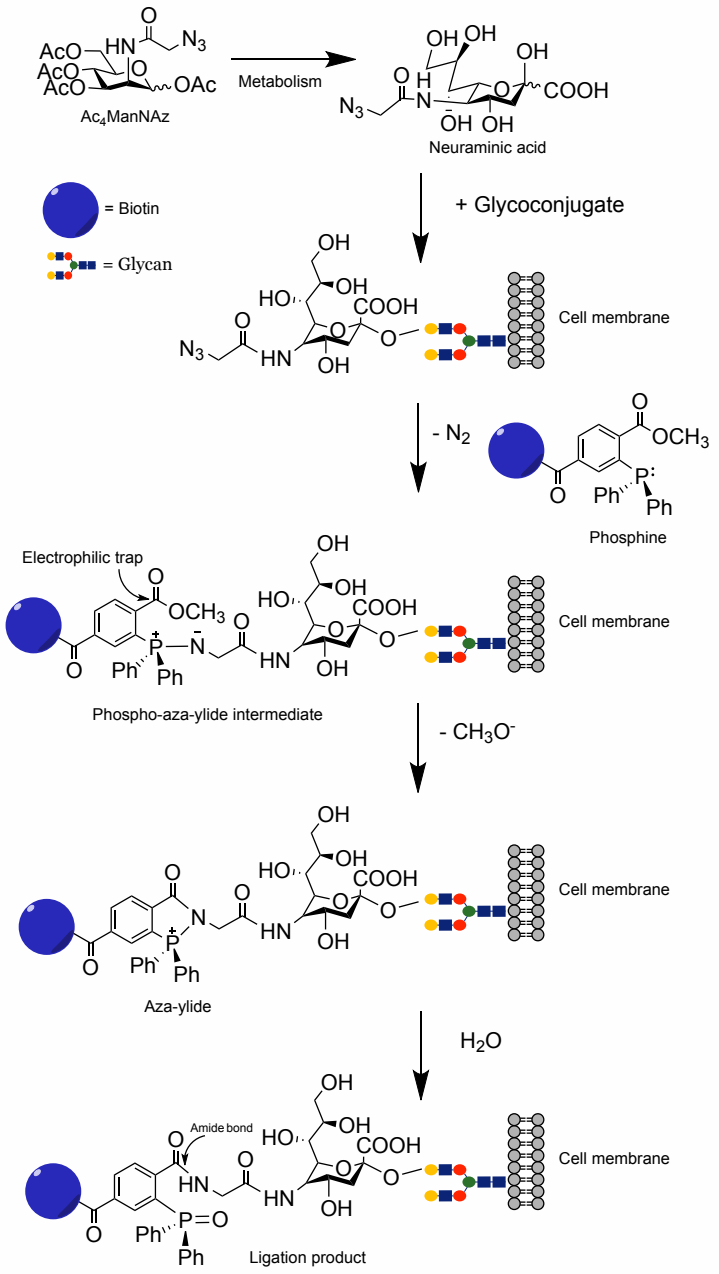
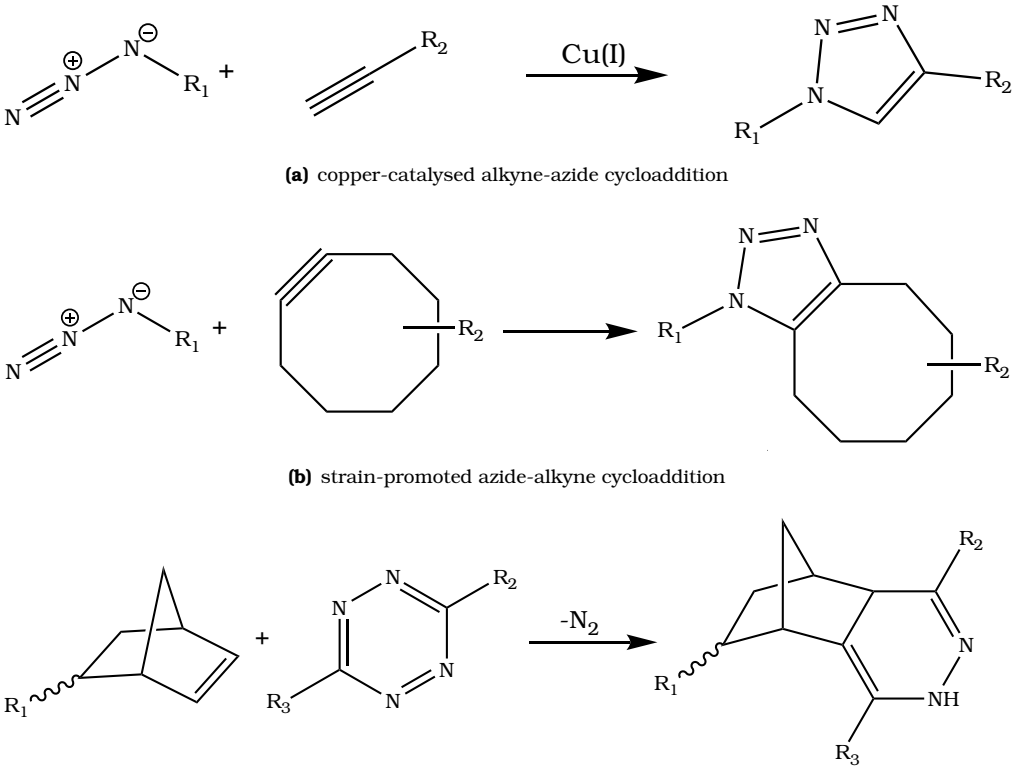


Figure 2.5: Metabolic incorporation of an azide-functionality into the glycoproteome of a cell. Metabolic conversion of ManNAz to SiaNAz and subsequent labelling with formation and trapping of the phospho-aza-ylid intermediate.



(a) copper-catalysed alkyne-azide cycloaddition

(b) strain-promoted azide-alkyne cycloaddition

(c) inverse Electron Demand Diels-Alder ligation

Figure 2.6: Bioorthogonal reactions. (a) CuAAC - copper-catalysed alkyne-azide cycloaddition. (b) SPAAC - strain-promoted azide-alkyne cycloaddition. (c) iEDDA - inverse Electron Demand Diels-Alder ligation.

cluding the need for copper. This so-called strain-promoted azide-alkyne cycloaddition (SPAAC), whilst not as fast as the CuAAC ($<10^{-2}$-1 $M^{-1}s^{-1}$), has been used successfully to label azides even in live zebra fish[171] and mice[172] (Reaction b, Figure 2.6). One disadvantage of the SPAAC ligation is that the strain in the alkyne also increases its side-reactivity profile, with reactivities towards thiols[173,174] and oxidative conditions having been reported[175]. One final class of bioorthogonal reactions that warrants mentioning is the family of inverse electron demand Diels-Alder (iEDDA) ligations (Reaction c, Figure 2.6). Its first use as a bioorthogonal ligation was reported by Fox and coworkers[77]. iEDDA reactions occur between electron-poor dienes with an electron-rich dienophile. The reactions can be extremely fast with rate constants of ~ 1-2 x 10^4-2000 $M^{-1}s^{-1}$ for the ligation having been reported[176]. It can also readily be made fluorogenic due to the tetrazine having a capacity to quench attached fluorophores until the point of ligation to the dienophile. This can provide an important additional advantage for biological experiments[177–179]. Many more bioorthogonal reactions have been reported, reviewed in Lang et al. 2014[176], and all have their uses and drawbacks for application in biological systems. Arguably, no truly bioorthogonal reaction has yet been developed, with regards to reagent stability, toxicity, and side-reactions. However, the approach is

still a useful addition to the biological tool kit.

2.4.2 Incorporating bioorthogonal groups into proteins

The first step of a bioorthogonal approach is the introduction of one of the ligation partners into a biomolecule in a cell. This has been done for the aforementioned carbohydrates[171], lipids[180], DNA[181], RNA[182] and most importantly for this thesis, proteins[183,184]. In order to introduce chemical functionalities co-translationally into a protein two approaches are available: bioorthogonal non-canonical amino acid tagging (BONCAT) and amber codon suppression.

BONCAT was first introduced by Tirrell and co-workers[185]. Their work, using methionine auxotroph (met^{aux}) *Escherichia coli* (*E. coli*) strains, allowed them to introduce unsaturated hydrocarbon-amino acids into protein backbones[186,187] and relies on the substitution of naturally occurring amino acids by isosteric bioorthogonal counterparts. After the discovery of the reactivity of some of these isosteres (i.e., the alkyne) in bioorthogonal ligation reactions they used the bacteria to produce recombinant proteins containing multiple alkynes or azide-containing side chains (introduced as the amino acids azidohomoalanine (Aha) or an alkyne (homopropargylglycine; Hpg), which then could be ligated using the Staudinger ligation[188], or the CuAAC[189] to produce recombinant proteins[2,185,186,190]. The approach was expanded to the incorporation of Met-analogues into non-auxotrophic bacteria[191], mammalian cell lines[192] or whole mice[193].

Through the introduction of mutated tRNA/tRNA synthase pairs selective for natural codons[194], the approach was also expanded to allow the replacement of other amino acids (i.e., tRNA-Phe, tRNA-Ala, tRNA-Ser[195]). The advantage of this approach is that each protein (including the recombinant proteins that can be purified) contains multiple copies of the bioorthogonal ligation handle, usually in reasonable protein yields. Downsides include the incomplete label incorporation, with some exception[190], as well as an effect on viability and growth of the target cell[191,196]. However, despite these limitations, this approach has recently been used to label a whole mouse proteome[197].

Amber codon suppression on the other hand, makes use of mutant aminoacyl transfer RNA synthases (aaRS) that can load tRNAs recognizing the amber stop codon with unnatural amino acids (UAAs)[198]. This permits the incorporation of bioorthogonal amino acids as a 21^{st} amino acid. In the first example of this approach, a tRNA/tRNA synthase from the archaebacteria *Methanocaldococcus jannaschii* was transformed into *E. coli*[198]. In the first examples combined with bioorthogonal chemistry, this tRNA/tRNA synthase pair was mutated to recognize p-azidophenylalanine, which was subsequently used to introduce a fluorophore into the protein backbone[199]. The identification of tRNA/tRNA-synthase pairs from the archaea *Methanosarcina barkeri* and *Methanosarcina mazei* that were capable of introducing pyrrolysine analogues further expanded the scope of this approach. This method was used to introduce azides[200], alkynes[201,202] and reactive alkenes[203] into proteins and ligate these selectively. Amber codon suppression allows, by and large, the introduction of only a single bioorthogonal group into a protein. The incorporation of multiple copies of an amino acid have been reported[204,205], however, at much reduced yields. Yet, the addition of this single amino acid in a native backbone is still very powerful[184].

2.4.3 Bioorthogonal chemistry during proteolytic degradation

As already mentioned above in Section 2.4 the classical reporter strategies such as genetic tagging with fluorescent proteins or the covalent modification of nucleophilic side chains are of limited compatibility with the study of proteolysis. Detectable fusion partners, for example, are only visible until they themselves are degraded, which during the process of antigen presentation, therefore can only allow the imaging of the parts of the process where this has not yet occurred [129,206]. Whereas this has yielded important information on the early steps of antigen processing, it has left a 'black box' regarding the fate of antigen between its partial degradation and the on-surface appearance of peptide. It may have even introduced bias regarding the potential routes of the antigen, as only those molecules, which remain visible have been used to devise hypotheses on the routing of antigen. Even the use of fluorophore modified antigens may be limiting in these regards. Van Elsland *et al.* presented the first use of bioorthogonal antigens as possible solution for studying degradation in professional APCs [1,207]. By incubating APCs with bioorthogonally labelled *E. coli*, they could study the degradation of these bacteria over time in the cell. An additional highlight of their work is the addition of the ultra-structural information of the phagosomal system of the used APCs using correlative light-electron microscopy (CLEM) [208]. In this thesis, BONCAT was used as an approach to produce bioorthogonal antigenic model proteins. In an approach similar to the approach described above, recombinant bioorthogonal proteins were produced and their degradative properties in the cell studied, as well as their ability to activate T cells after processing.